DOCUMENT RESUME

ED 278 631 SP 028 520

AUTHOR Sia, Archibald P.; Sydnor, Darlean
TITLE Promoting Excellence: Some Classroom Evaluation Alternatives.
PUB DATE 17 Feb 87
NOTE 20p.; Paper presented at the Annual Meeting of the Association of Teacher Educators (67th, Houston, TX, February 15-18, 1987). Table 1 contains small, broken type.
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Classroom Observation Techniques; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Rating Scales; School Effectiveness; *Student Evaluation; *Testing

ABSTRACT

Following a brief overview of literature on excellence in the classroom, the shortcomings of objective tests as means for evaluating classroom excellence are discussed as well as the need for evaluation alternatives. Descriptions are given of evaluation alternatives, some of which focus upon reading and writing achievement: (1) observation of actual task performance; (2) documentation of student work samples; (3) use of rating scales; (4) assessment of writing samples; (5) student workbooks; (6) student debriefing after a lesson; (7) role playing and simulation games; (8) performance checklists; (9) case studies and anecdotal records; (10) interviews; and (11) questionnaires and opinionnaires. A three-page list of references and tables that illustrate some of the evaluation methods conclude the document. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED278631

PROMOTING EXCELLENCE: SOME CLASSROOM EVALUATION ALTERNATIVES

by

Archibald P. Sia, Ph.D.
Assistant Professor
Rockhurst College
Kansas City, MO

and

Darlean Sydnor, Ph.D.
Assistant Professor
Pacific Lutheran University
Tacoma, WA

A Paper Presented for the 1987 Annual ATE Conference
Houston, Texas
February 17, 1987

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. P. Sia

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SP 028 520

BEST COPY AVAILABLE

PROMOTING EXCELLENCE: SOME CLASSROOM EVALUATION ALTERNATIVES

Promoting excellence among our students is a goal of education. However, recent documents, like A Nation at Risk (1983), the Holmes Group report (1986), A Place Called School (1984), and others charge that our schools hardly promote excellence. No matter how educational excellence is conceived, this indictment of our schools necessitates serious attention. If the goal of promoting student excellence is to be realized, there is a need to have a clear conception of what constitutes excellence, whether excellence can be quantified in objective test terminology, and then to explore what alternatives are available to us in assessing our goals toward excellence.

EXCELLENCE: VARIED VIEWS

Defining excellence is no easy task. It may be defined at the student level, the school level, the community level, and at a national, or even global level. In this paper, excellence will be viewed from the level of the individual student and the ways of evaluating individual progress towards excellence. The National Commission on Excellence in Education (1983) proposes a definition of student excellence based on each student's performance on the boundary of individual ability in ways that test and push back personal limits, in school and in the workplace.

If we accept the above definition of excellence, it would seem that the reforms that are currently taking place do not strive for educational excellence but educational adequacy (Johnson, 1982). The emphasis on course requirements and more testing as measures of excellence is one example of this claim (Quellmalz, 1985).

To conceive of excellence as test performance and credit accumulation has led reformers to think in terms of standards of excellence, standards which Duke (1985) believes often resembled minimum competencies. Johnson (1982) questions the use of standards of excellence, for he maintains that standards imply standardization which could be considered the absence of excellence. Instead of standards of excellence, therefore, he suggests that we use criteria of excellence, models of excellence, even visions of excellence. Johnson's conception of excellence makes us aware of the danger of supporting a single image of excellence - one tied to credit accumulation and performance - which could threaten the very assumptions on which a pluralistic, egalitarian society is based (Duke, 1985).

Gardner (1983) supports Johnson's view of varied images of excellence through his theory of multiple intelligences. Identifying seven separate intelligences, he is highly critical of the heavy - and at times exclusive - emphasis on logico-mathematical intelligences in U.S. schools.

As there are different ways of looking at educational excellence, trying to capture it as a single entity that can be measured is counter-productive. Reifying excellence this way may lead reformers

to rely heavily on standard evaluation techniques as ways of monitoring excellence (Gould, 1981). Excellence defies upper limit and it simply cannot be captured by objective tests.

SHORTCOMINGS OF OBJECTIVE TESTS

The exclusive use of objective tests in evaluating pupil progress towards educational excellence has been under strong attack. The criticisms include the inadequacy of objective tests in measuring problem solving and other higher order skills (Fredericksen, 1984; Quellmalz, 1985; Stedman, 1985), their potential to harm good reading skills and writing ability (Meier, 1984; Harrison, 1980; Suhor, 1985), the possibility of their thwarting creativity and aesthetic sensitivities (Swanson & Zhang, 1987; Cuban, 1983), and their traditional racism and sexual bias (Gould, 1981).

Fredericksen (1984), in a review of research on conditions that encourage problem solving, failed to find that increased testing was critical. He observes that problem solving factors, e.g., adequate time for incubation of ideas and opportunity to practice higher-order cognitive skills in permissive environments in which failure is acceptable, are missing.

This finding is shared by Quellmalz (1985) who states that subject matter achievement tests present few items that measure higher order skills, and those items tend to require fractionalized reasoning skills. Even specialized thinking skills limit reasoning to tasks that seek one right answer and allow no opportunity or requirement for explanation.

Stedman (1985) states that all the talk about test score declines and getting back to the basics tends to obscure the long standing failure of U.S. schools to teach higher order skills and to reach the lower third of students. He further adds that the recent focus on testing is already having serious negative consequences. Many teachers are returning to the old fashioned method of whole group instruction, combining lecture, recitation, and seatwork - a strategy that Cuban (1983) believes goes far beyond what the research suggests.

Another negative consequence of the focus on testing is that it has been found to harm the development of good reading skills (Meier, 1984). In inner city schools attended mainly by low-income children, Meier notes the prevalence of the program scripts based on behavior modification techniques, reading kits consisted of unrelated paragraphs followed by multiple choice questions and reams of ditto sheets. The problems worsen and spread as pressure for accountability grows and test score rehabilitation is made the central goal of education.

Objective testing results in more emphasis on the study of basic composition skills rather than written composition per se - a finding that is consistent with the widespread testimony that skills-based testing reduces the actual teaching of writing (Harrison, 1980).

Suhor (1985) confirms Harrison's findings that both standardized, norm-referenced tests and criterion-referenced competency tests discourage writing instruction. He states that such tests fragment and trivialize instruction while tests that include writing samples generally result in more teaching of writing. Swanson and Zhang (1987) find the examination system to have contributed to an overemphasis on book learning and to have developed respect for authority to the detriment of creativity, inquisitiveness, and an enterprising spirit. They state that reliance on a test score alone to determine academic advancement ignores other important indicators of academic success.

Cuban (1983) criticizes the single-minded quest for higher test scores for narrowing the school agenda to content that is easily measured. This generally limits the thrust of the program to aspects of the cognitive domain which are easily quantified. At this point, one must wonder any commitment or concern for the other goal commitments of schooling. An emphais on testing mainly on the cognitive domain leads to less attention to sharing, developing self-esteem and acquiring aesthetic sensibilities.

Gould (1981), in his book Mismeasure of Man, notes the implicit and explicit racial and sexual biases in the early development of testing and measurement.

Callahan (1986), commenting on gifted education, states that the traditional assessment of inputs on the basis of multiple choice items severely limits the kinds of evaluation evidence we can gather and transmit to interested parties. Anrig (1986) shares this view of test limitations and the restricted range of qualities they can measure.

Exploring the use and misuse of tests, Resnick and Resnick (1985) show evidence from Pittsburg that minimum competency testing encourages teachers to focus only on the minimum skills to be tested. Tyler (1978) concurs with this finding. He reports that some teachers actually believed that Florida law required them to limit the curriculum to the minimum competencies required by the tests. They interpreted the emphasis on basic skills as a mandate to give maximum attention to route drill.

From the preceding data, we can infer that the use of conventional measurement instrument as the primary mechanism for monitoring educational excellence is questionable, and, in fact, may even impede excellence. Testing may be fine for insuring educational adequacy, but it seems less valuable as a way of measuring the push beyond minimum expectations. The temptation to think that everything worthwhile can be quantified - using formal tests of achievement - should be actively resisted. Alternative evaluation measures must be explored.

THE NEED FOR EVALUATION ALTERNATIVES

Worthen and Sanders (1973) agree that all of our measurements are fallible. They all have their weakness. Therefore, we should try as much as possible to index in a variety of ways the things in which we are interested. It will be "more instructive and more convincing if our measures have compensatory strengths and unshared weaknesses".

Callahan (1986) echoes similar observation. He believes that evaluation questions are limited because of lack of alternative data sources such as objective and subjective product assessment. He states that the tradition of assessing impacts on the basis of multiple-choice test items severely limits the kinds of evaluation evidence we can gather and transmit to interested parties.

Evidence for evaluation may be provided through a sound evaluation data collection (Shapiro et al, 1986), which should not be limited to written form. Students' oral behavior can be important, and teachers can also directly guage a child's progress if a systematic program of informed assessment is in place. At times such approaches should be central to the evaluation program.

The need for alternative measures is shared by Engel (1975). He states that conventional testing methods have proved not only damaging and unreliable but also grossly inadequate. The focus of the tests, no matter what their merits, is too narrow to provide grounds for reasonable general conclusions about individuals' progress. Berlak (1985) echoes this need for alternative measures when he states that resources must be diverted to the development and use of more subtle, sensitive and educationally useful forms of assessment.

TOWARD STUDENT EXCELLENCE: SOME EVALUATION ALTERNATIVES

Perhaps the most comprehensive evaluation alternative to objective testing is through documentation. This process is a selection of documents put together in order to portray an event as clearly and honestly as possible, with its significant elements represented (Engel, 1975). The documents consist of both sampling and summaries, presented in various forms such as accounts, diagrams, graphs, mind maps, photographs, drawings, schedules, lists, formal assessments, transcribed tapes, etc. Documentation, as an alternative, offers a better possibility for obtaining useful evaluation data since it can be better correlated with the goals and content of the program.

What follows are some evaluation alternatives which can be incorporated as parts of the documentation process.

Observation methods. This alternative includes direct observation of actual task performance or examination of a sample of completed work. Subjects like music, physical education, and home economics may easily make use of this method. Der-Breme (1983) states that in 1983, the Center for the Study of Evaluation, using a large, randomly selected national population, discovered that teachers and principals have more confidence in such data as teacher observation and the results of teacher-made tests than they do in the results of standardized tests.

Cryan (1986) gives examples of observation methods, such as: (1) specimen description, where the observer identifies a situation and records all elements within that situation, (2) time sampling, where previously defined and targeted behaviors are coded within discrete time units, and (3) event sampling or anecdotal records, where critical incident or behaviors previously determined as targeted are recorded.

Worthen and Sanders (1973) note that simple and contrived observation may serve as evaluation tools. The latter uses hardware devices for observation, e.g., VTR and audiotapes. These resources enable the teacher to develop his/her skills in the systematic observation and recording of student behavior in non-test contexts (Shapiro et al, 1986).

Other ways of direct observation involves student performances. This alternative monitors activities in progress, like playing a song, drilling a hole, giving a speech, shooting a basket, teaching a lesson, administering CPR, or dismounting from parallel bars. Patton (1981) calls this alternative demonstrations.

Still another direct observation alternative is documenting student work samples, whereby products of student activities are monitored, e.g., plan for a solar heater, papier mache model of landscape, a birdhouse, bookcase, letter to editor, etc. Criteria in rating performances may be applied in judging work samples. In practice, teachers can observe only that which is possible within limitations of class size, time, location, and nature of task.

Other Alternatives to Evaluating Work Samples. Student work samples may further be self-evaluated by the use of rating scales. This method may take three forms: numerical, graphic, or comparative. A numerical scale has numbers to indicate the degree to which a particular characteristic is present whereas a graphic scale is a rating made by checking along a line. A comparative scale provides samples of a product to be evaluated, usually at 5 to 7 levels of quality.

Cryan (1986) notes that samples of children's work may be used as direct evidence of progress and if dated when collected the samples may be interpreted later and by persons other than the teacher, perhaps even by the student himself/herself. Evidence is provided by McCaig (1977). His report shows marked improvement of grades 1-6 writing ability and attributed it to the centralized evaluation of student writing by means of a writing sample.

Although there is only a small amount of empirical data available on the testing of writing, Suhor (1985) notes that the use of writing samples in evaluating writing per se supports the overwhelming testimony of practitioners. The data also indicate the use of other alternatives to writing assessment.

Alternatives to Writing Assessment. Students' writing ability may be assessed directly through the following alternatives: (1) holistic, (2) analytical, (3) primary trait, (4) writing mechanics, and (5) t-unit analysis (National Council of Teachers of English, 1980). Table 1 shows a brief summary of each of these writing alternatives.

INSERT TABLE 1 ABOUT HERE

From the above table, we can see the general capabilities, reliability, amount of preparation, scoring time, sample size, and classroom use of the five alternatives. If, for example, our objective is to assess students' ability to use mechanics properly, the writing mechanics alternative may be the best one to use. If our objective is to assess students' ability to report overall writing skill, then the holistic approach may be the relevant alternative.

Still another alternative available for evaluating reading and writing is the use of workbooks. Lapointe (1986) states that 95% of teachers use this approach. When used with minorities, e.g., blacks and Hispanics, slight increase in reading ability was noted. The above alternatives, from student samples to workbooks, involve to a large extent the use of records and observational techniques. Evaluating student progress through the use of these methods may necessitate that we entertain the following questions: Who is going to use the information? What do they need to know and do we provide it? For what purposes will they use the information? What should be the proportion of subjective to objective data? Is there a case of showing pupils' own work?

Our record keeping will be greatly aided by videotaping. Beegle and Brandt (1973) state that videotaping is helpful in checking observer reliability in controlled observation areas such as specific description and event sampling. This alternative has not only allowed for multiple viewing of observation periods, but also, because of its applicability both as a model and a purveyor of self-feedback, has aided in the training of human observers.

Debriefing. Another alternative for classroom evaluation is debriefing (Duke, 1977). This involves students' own reconstruction of prior learning experiences. The teacher asks the students to express in their own words what they have learned after a lesson, unit, or course. This alternative puts on the students the ability to take responsibility for organizing, retaining, and making sense of what they are exposed to in classes.

Role Play and Simulation Games. Another powerful and proven technique to practice new behaviors, skills, and approaches is role playing (Patton, 1981). This alternative also helps students to understand perspectives and situations other than their own and develop empathetic attitude. Simulation gaming allows participants to take roles which are representatives of roles in the real world. Simulation games are ways of creating microcosms of real world situations in a classroom. As it is usually problem-based, it is helpful in developing interdisciplinary approaches to learning and

use of social skills. This alternative was effectively utilized in an environmental education class where students assume roles pertaining to their views on population issues.

Checklists. This alternative is intended to structure observation, focusing it upon significant aspects of children's behavior and at the same time providing a framework for interpreting the observation in terms of progress towards its goal (Harlen, 1978). Very easy to use, a checklist provides immediate descriptive information allowing for on the spot evaluation and rapid feedback about the adequacy of performance or production (Cryan, 1986).

Checklists are among the most useful of the informal methods when the performance involves a large number of specific acts or the product consists of a great number of dimensions. Some examples monitor sentence-processing skills, dictionary and reference skills, critical thinking skills, map-reading characteristics, research paper characteristics, social participation scale, work assignment progress list. An example of a checklist which addresses work assignment progress list is on the use of mathematics manipulatives, as shown in Table 2a. An observation checklist on the use of a clinical method in science is shown in Table 2b.

INSERT TABLES 2a & b ABOUT HERE

Case Studies and Anecdotal Records. Case studies involve following up students progress along the dimensions of interests. Anecdotal records are written report of specific incidents. Both approaches are used in virtually every subject area. Their success in special education classes have been recorded (Engel, 1975).

Interviews. This alternative is roughly akin to the live administration of a questionnaire. Popham (1975) notes several advantages of interviews. This process can put the subject at ease thereby securing more candid responses. Furthermore, the interviewer can follow up responses, probing into students' thinking. Interviews may be structured or less structured to give the student more freedom. This could provide valuable information and is more suitable for use in almost any type of classroom or other setting. Cryan (1986) suggests that the interview method be used with other forms of evaluation to supplement or extend the meaning of the data. Table 3 shows a clinical interview method used in a science/math class.

INSERT TABLE 3 ABOUT HERE

Questionnaires and Opinionnaires. Though questionnaires cannot probe in the way that might be accomplished by an interview, this alternative has some distinct advantages. These include: (1) time factor and (2) its ability to assess larger number of students thereby improving the reliability of information (Cryan, 1986).

An example of the use of an opionionnaire is reported by Harlen (1978). A pupil's opinion survey on a course is developed whereby a pupil is asked questions about reactions to the course, the activities which he/she finds hard, easy, most enjoyable, what has been learned most, and how pupils have perceived the demands of the material.

EXCELLENCE: THE UNENDING QUEST

Certainly the above alternatives to classroom evaluation are already in use, especially by those teachers who strive for educational excellence. As the concept of excellence is applied to the development of the child's fullest potential, more dedicated teachers will unfold varied evaluation alternatives. The challenge is for each teacher to find a philosophically appealing and educationally effective balance between common experiences and cultural diversity, between a supportive atmosphere and visions of excellence, and the transmission of uplifting knowledge. In this process tests can play only a limited role (Stedman, 1985). In the end, it is the pursuit - rather than achievement - of excellence that matters most. Placing value on the pursuit of excellence permits more students to feel good about themselves and to appreciate the accomplishments of others (Duke, 1985). A teacher who subscribes to this belief will strive to foster educational excellence for everyone and will utilize varied evaluation alternatives in the process.

REFERENCES

Anderson, S.B. et al (1975). Encyclopedia of educational evaluation. San Francisco, CA: Jossey-Bass Publishers.

Anrig, G. (1986). Teacher education and teacher testing: The rush to mandate. Phi Delta Kappan, 67(6), 447-451.

Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. Journal of Teacher Education, 36(4), 46-49.

Beegle, C.W. & Brandt, R.M. (1973). Observational methods in the classroom. Washington, D.C.: ASCD.

Berlak, H. (1985). Testing in a democracy. Excerpted from Equity and excellence: Toward an agenda for school reform. St. Louis, MO: The Public Information Network.

Brooks, D. & VanCleaf, D.W. (1982). Pupil evaluation in the classroom. Lanham, MD: University Press of America.

Callahan, C.M. (1983). Issues in evaluating progress for the gifted. Gifted Child Quarterly, 27(1), 3-7.

Callahan, C.M. (1986). Asking the right question: The central issue in evaluating programs for the gifted and talented. Gifted Child Quarterly, 30(1), 38-42.

Cryan, J.R. (1986). Evaluation: Plague or promise? Childhood Education, 62(5), 344-350.

Cuban, L. (1983). Effective schools: A friendly but cautionary note. Phi Delta Kappan, 64(10), 695-696.

Der-Breme, D. & Herman, J.L. (1983). Testing in schools: A national profile. Los Angeles: Center for the Study of Evaluation, UCLA.

Dressel, P.L. (1976). Handbook of academic evaluation. San Francisco: Jossey-Bass Publishers.

Duke, D.L. (1985). What is the nature of educational excellence and should we try to measure it? Phi Delta Kappan, 66(10), 571-674.

Duke, D.L. (1977). Debriefing: A tool for curriculum research and course improvement. Journal of Curriculum Studies. vol. 9, 157-163.

Engel, B.S. (1975). A handbook of documentation. Grand Forks, ND: University of North Dakota Press.

Fredericksen, N. Implications of cognitive theory for instruction in problem solving. Review of Educational Research, Fall 1984, 363-408.

Gardner, H. (1983). Frames of mind: A theory of multiple intelligences. New York: Basic Books.

Goodlad, J.I. (1984). A place called school: Prospects for the future. New York: McGraw-Hill Book Company.

Gould, S.J. (1981). The mismeasure of man. New York: W.W. Norton and Company.

Harlen, W. (ed.) (1978). Evaluation and the teacher's role. London: Macmillen Education, Ltd.

Harrison, M.J.C. (1980). Florida high school language arts curriculum changes resulting from the Florida state student assessment test: Status and impact of changes. Doctoral dissertation, Florida State University.

Holmes Group Report. (1986). Tomorrow's Teachers. 501 Erickson Hall, East Lansing, MI.

Johnson, M. (1982). The continuing quest for educational quality. School of Education, State University of New York at Albany.

Lapointe, A. (1986). The state of instruction in reading and writing in U.S. elementary schools. Phi Delta Kappan, 68(2) 135-138.

McCaig, R. What research and evaluation tells us about teaching written expression in the elementary school. In Weaver, C & Donma, R, eds. (1977). The language arts teacher in action. Kalamazoo: Western Michigan University, 46-56.

Meier, D. (1981). Why reading tests don't test reading. Dissent. Fall 1981, 457-466.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, D.C.: U.S. Government Printing Office.

National Council of Teachers of English. (1980) Direct writing assessment. Urban, Ill.

Patton, M.Q. (1975). Alternative evaluation research paradigm. Grand Forks, ND: University of North Dakota Press.

Patton, M.Q. (1981). Creative evaluation. Beverly Hills, CA: Sage Publication.

Poggio, J. & Glasnapp, D. (1983). An assessment of the impact of minimum competency testing in Kansas. Lawrence: Department of Educational Psycology and Research. School of Education, University of Kansas.

Popham, W.J. (1975). Educational evaluation. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Quellmalz, E.S. (1985). Needed: better methods for testing higher-order thinking skills. Educational Leadership, 43(2), 29-35.

Renzulli, J.S. & Reis, S.M. (1985). The school-wide enrichment model. Mansfield Center, CT: Creative Learning Press.

Resnick, D. & Resnick, L. (1985). Standards, curriculum, and performance: Historical and comparative perspective. Educational Researcher, vol. 14, 5-20.

Shapiro, P., Lukasevich, A. & Shapiro, P. (1986). Student evaluation: a change of focus. Education Canada, 26(2), 26-29.

Stedman, L.C. & Kaestte, C.F. (1985). The test score decline is over: Now what? Phi Delta Kappan, 67(3), 204-210.

Suhor, C. (1985). Objective tests and writing samples: How do they affect instruction in composition? Phi Delta Kappan, 66(9), 635-639.

Swanson, A.D. & Zhang, Z. (1987). Education reform in China. Phi Delta Kappan, 68(5), 373-378.

Tyler, R. et. al. (1978) The Florida accountability program: An evaluation of its educational soundness and implementation. Washington, D.C.: National Education Association.

Worthen, B.L. & Sanders, J.R. (1973). Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones Publishing Co.

TABLE 1

ALTERNATIVES TO WRITING ASSESSMENT

DESCRIPTOR	HOLISTIC	ANALYTICAL	PRIMARY TRAIT	WRITING MECHANICS	T-UNIT ANALYSIS
GENERAL CAPABILITIES	Comprehensive, general picture of student performance; writing viewed as a unified coherent whole. Applicable to any writing task.	Thorough trait by trait analysis of writing; provides comprehensive picture of performance *if* enough traits are analyzed; traits are those important to *any* piece of writing in *any* situation (e.g., organization, wording, mechanics).	Highly focused analysis of situation-specific primary trait (and possibly secondary traits); provides specific information on a narrowly defined writing task (e.g., ability to recount details in chronological order).	Can provide either a general or a specific profile of the student's ability to use mechanics properly.	Provides a measure of syntactical sophistication.
RELIABILITY	High reliability if standards are carefully established and raters are carefully trained.	High reliability if criteria and standards are well defined, and careful training is conducted.	High reliability if criteria and standards are well defined, and careful training is conducted.	High reliability if given sufficient training time and authoritative, complete, acceptable guidelines (e.g., an English handbook).	High reliability provided trained and experienced raters are used.
PREPARATION TIME	Up to one day per item to identify range finder (model) papers; up to one-half day to train readers using 4-point scale; full day to train with 8-point scale.	One full day to identify traits; one day per trait to develop scoring criteria (unless traits and criteria are borrowed from another source); one to two days to review results of pilot test and refine traits or criteria as necessary; one-half day to train raters.	One full day to identify traits; one day per trait to develop scoring criteria (unless traits and criteria are borrowed from another source); one to two days to review results of pilot test and refine traits or criteria as necessary; one-half day to train raters.	One to two days to set up a scoring system (unless borrowed from another source). Minimum of one day to internalize the scoring system and practice scoring.	Half day to full day, depending on raters previous experience.
READERS	Qualified language arts personnel recommended; high reliability can be achieved with non-language arts readers given sufficient training.	Qualified language arts personnel recommended.	Qualified language arts personnel recommended; non-language arts staff may be able to score some traits.	Qualified language arts personnel recommended.	Raters *must* be experienced language arts personnel, preferably those already familiar with the concept of T-unit analysis.
SCORING TIME	One to two minutes per paper (experienced readers may read faster).	One to two minutes per paper per trait.	One to two minutes per paper per trait.	Five minutes or more per paper, depending on number of criteria.	Varies greatly, depending on raters skill.
CLASSROOM USE	May be adapted for use in class.	May be adapted for use in class.	May be adapted for use in class.	May be adapted for use in class.	May be adapted for use in class.
REPORTING	Allows reporting on students' overall writing skill.	Allows reporting of student performance on wide range of generalizable traits (i.e., the qualities considered important to all good writing).	Allows reporting of student performance on one or more situation-specific traits important to a particular task.	Allows reporting of group or individual data on students general strengths or weaknesses in mechanics.	Allows group or individual reporting on syntactical sophistication.
GROUP/ SAMPLE SIZE*	Primarily usable with a larger sample; with a small sample, responses may be difficult to scale.	Best with smaller samples; extensive scoring time *may* make costs prohibitive with larger groups.	Generally more cost-effective with smaller samples, depending on the number of traits to be scored (with one trait, sample size is not an issue).	Best with smaller samples; extensive scoring time *may* make costs prohibitive with larger groups.	Best with smaller samples; extensive scoring time *may* make costs prohibitive with larger groups.

TABLE 2a

A MATH LAB COMPETENCY CHECKLIST

Name: _____ Ed.112: Teaching Elementary Math

The mathematics laboratory is designed for a discovery approach to learning. Try to learn concepts by using the aids as though you were a child learning the concepts for th first time. Begin by experimenting with the aid to discover its uses. Continue experimenting with the aid. Your textbook (various chapters and appendix) and other laboratory resources are available for your use in learning the aids. Be thoroughly familiar with the aids so that as an elementary school teacher you will be able to use the aids with your students. The teaching assistants are available to give competency checks. They will also help you locate materials related to the concept or aid you are learning. Return this list to your instructor by the due date stated.

Manipulative Aids	Date - Signature
1. Number Balance	_____
2. Number Line	_____
3. Multibase Blocks	_____
4. Attribute Shapes	_____
5. Cuisenaire Rods	_____
6. Mira	_____
7. Geoboard	_____
8. Fraction Kit	_____
9. Tangrams	_____
10. Abacus (Japanese)	_____
11. Minicomputer	_____

TABLE 2b

A SCIENCE OBSERVATION CHECKLIST

CLINICAL METHOD OBSERVATION CHECKLIST

Name of Experimenter (E) ..... Date .....

Name of Subject (S) .....

Name of Observer .....

Check the response you feel is best based on your observation. Give examples where space is provided.

Introduction

1. (a) E's efforts to make the S feel at ease were ....

Very Good	Good	Satis-factory	In need of improvement

(b) What evidence do you have for your answer?

.....

.....

2. Did E. explain the purpose of the interview to S?

Did S. understand?

What evidence do you have for your answer?

Yes	No

.....

.....

Questioning

3. The language E. used to communicate with S. was

Very Ap propriate	Satis factory	Often inap propriate

Give examples (e.g. use of familiar words, simple sentences. . .)

.....

.....

4. The non-verbal behavior of E. enabled S. to get cues to the answers required. . . . (e.g. posture of E. position of E. facial expressions. . .)

Often	Sometimes	Hardly ever

5. The verbal behavior of E. enabled S. to get cues to the answers required(e.g. leading questions. praise of "correct" answers

Often	Sometimes	Hardly ever

6. E. used a flexible approach to questioning (i.e. E. varied the prepared protocol questions).

Always	Mostly	Hardly ever varied

Yes	No

7. Was S. given enough time to answer the questions?

8. Was S. encouraged to elaborate on his answers? e.g. asked to explain, justify

. . . . . .

9. Did E. use the terms which S. used spontaneously?

10. E. made use of counter-suggestions.

Often	Sometimes	Never

Give example.

.....

.....

TABLE 3

A SAMPLE CLINICAL INTERVIEW IN SCIENCE

UNDERSTANDING DISPLACEMENT OF VOLUME

Recording Sheet

Stage: .....
Reasoning .....

Experimenter: Lilia Reyes
Recorder: Buena Lozada
Date: October 11, 1978
School: Glugor Elementary

Name: Leona Boey
Age: 10 Years 2 Months
Sex: Female
Grade: IV

Experimenter's Actions/ Questions	Child's Actions/Opinions
Constant Volume, Change in Shape of Sinker	
Ask the child to fill up the can to the level of the overflow point.	
Immerse a ball of plasticine in the overflow can and allow the overflow to collect in a glass container.	
"Can you tell me what has happened?"	The water came out of the spout.
Roll the ball of plasticine into the shape of an egg.	
"If I put this egg-shaped plasticine into the water in the can, would the *same* amount of water overflow? or *more?* or less than before?	The same as the amount of the first overflow.
"Why do you say so?"	The egg has the same amount of plasticine as the ball.
Roll the plasticine into the shape of a sausage.	
"If I put this sausage-shaped plasticine into the can would *more* or *less* water overflow than on the first occasion?"	It would be the same. Although it's a ball or a sausage it has the same amount of plasticine and so it won't make any difference at all.
Countersuggestion.	
Another girl told me there would be more overflow than on the first occasion, do you think she was right?	No.
What would you do to convince her that she was wrong?	I would put an egg-shaped plasticine into the container and measure the amount of water overflow. I would roll the egg-shaped plasticine into the shape of the sausage. Put it into the can: compare the amount of the second overflow with the first.